PRACE INSTYTUTU BADANIA KONIUNKTUR GOSPODARCZYCH I CEN

STUDIES AND REPORTS OF THE POLISH INSTITUTE FOR ECONOMIC RESEARCH

Redaktor / Editor Prof. Edward Lipiński

Warszawa 1939 Nowa seria Nr 2

METODA „LINIOWA" ELIMINOWANIA WAHAŃ SEZONOWYCH

THE „LINEAR" METHOD OF ELIMINATING SEASONAL FLUCTUATIONS[1])

(Opracował J. Wiśniewski)

Doświadczenie Instytutu Badania Koniunktur Gospodarczych i Cen, jak niewątpliwie i innych instytucji tego rodzaju, zmusiło do zwrócenia uwagi na zmiany, jakim podlega kształt i rozpiętość wahań sezonowych. Szeregi statystyczne, z których wyeliminowano sezonowość za pomocą jakichś metod rachunkowych, „powinny" wykazywać tylko zmiany koniunkturalne i trendowe z jednej strony, a przypadkowe — z drugiej. Jednakże obserwacja takich szeregów statystycznych, szczególnie w latach kryzysowych, pozwoliła dostrzec właśnie coś w rodzaju wahań sezonowych. Zjawisku temu nadano początkowo miano „pseudo-sezonowości", gdy zaś stwierdzono jego systematyczność, usiłowano je zrównoważyć przez obliczanie układów współczynników sezonowości oddzielnie dla lat dobrej koniunktury, oddzielnie dla lat złej koniunktury. Jakie są przyczyny występowania zjawiska zmiennej rozpiętości wahań sezonowych w zależności od koniunktury, nie będziemy się tu zastanawiali, odsyłając czytelnika do prac dawniejszych[2]). Na tym miejscu zastanowimy się nad techniczno - statystyczną metodą eliminowania wpływu zmiennej sezonowości.

Schematycznie można by podzielić zmiany sezonowości (nie zmiany czyli wahania sezonowe) na dwie kategorie: strukturalne i koniunkturalne. Zmiany strukturalne mogą zachodzić szybciej lub wolniej, za wspólną ich cechę można uważać ich nieodwracalność (gdyby nawet zmiany strukturalne mogły w dalszej przyszłości zmienić kierunek, to jednak ograniczona perspektywa, jaką rozporządzamy, skłania nas do traktowania ich jako nieodwracalnych). Zmiany koniunkturalne natomiast traktujemy jako odwracalne. Dla wyeliminowania pierwszych możemy się posługiwać bądź to oddzielnym obliczaniem układów współczynników sezonowości dla poszczególnych okresów czasu, gdy zmiany sezonowości zachodzą dość raptownie, bądź też — gdy zmiany te są stopniowe — jedną z metod zaproponowanych przez pisarzy amerykańskich[3]). Metoda, którą opiszemy poniżej, jest dostosowana do zmian sezonowości, uzależnionych od zmian koniunktury. Jej myślą przewodnią jest ustalenie systemu współczynników sezonowości dla każdego poziomu koniunktury, a raczej dla każdego poziomu wartości „normalnych", będących wypadkową trendu i koniunktury. Wydawało się najprostszym przyjąć liniowy schemat zależności sezonowości od koniunktury; rachunkowo zresztą i ten „najprostszy" schemat wymaga sporo pracy. Być może, iż w

[1]) The English version of this paper was published in „Economic Studies", III, Kraków, 1936, pp. 55 — 68. Cfr. also „Econometrica", 1937, pp. 260 — 262.

[2]) J. Wiśniewski „Współzależność między wahaniami sezonowymi i koniunkturalnymi", Prace Instytutu Badania Koniunktur Gospodarczych i Cen zesz. 2 z r. 1933, str. 3—15. — J. Zagórski „Przyczyny zmian przebiegu sezonowości w statystycznych szeregach gospodarczych" tamże, zesz. 1 z r. 1935, str. 1—8.

[3]) King, Crum, Gressens i in. Cyt. u Landaua „Wahania sezonowe i metody ich mierzenia", Koniunktura Gospodarcza zesz. 1 z r. 1928, str. 67.

przyszłości nasunie się potrzeba uzupełnienia a więc skomplikowania tego schematu, np. przez wzięcie pod rozwagę wyższych potęg wartości „normalnych", albo też ich przyrostów (ta ostatnia alternatywa dałaby nam inne układy sezonowości dla lat polepszającej się, a inne dla lat pogarszającej się koniunktury).

Liniowy schemat sezonowości można przedstawić graficznie jak następuje. Na osi odciętych odkładamy wartości normalne *(a)*, na osi rzędnych wartości zmiennej rzeczywiście zaobserwowane *(x)*, po czym zaznaczamy punkty, odpowiadające kolejnym obserwacjom w ciągu n lat dla tego samego miesiąca. Stwierdzamy, że punkty te leżą w przybliżeniu na prostej, którą wyznaczamy przy pomocy metod opisanych poniżej. Jeżeli wybierzemy teraz jakąś określoną wartość a np. a_1, to możemy powiedzieć, że przy $a = a_1$ przeciętna wartość x wynosi x_1. Przypuśćmy teraz sytuację taką, że nie wiemy, jaka jest wartość a, natomiast mamy zaobserwowaną wartość x, np. x_2. Postępujemy wówczas odwrotnie, mianowicie znajdujemy, jaka wartość a odpowiada wartości x_2 na naszej prostej; jest to t. zw. eliminowanie sezonowości. Wyprowadzenie potrzebnych wzorów przedstawimy poniżej[1]).

Oznaczmy przez w wartość składnika sezonowego, przez a normalną wartość zmiennej[2]), przez x zaobserwowaną wartość zmiennej, wreszcie przez e wartość składnika przypadkowego. Wtedy tożsamościowo

$$x = a + w + e \qquad (1)$$

dla dowolnego momentu. Hipotezie stałej sezonowości addytywnej odpowiada zależność

$$w_i = \text{const.} \qquad i = 1, 2, \ldots 12 \qquad (2)$$

dla danego miesiąca i. Hipotezie stałej sezonowości multyplikatywnej odpowiada zależność

$$w_i = p_i\, a \qquad i = 1, 2, \ldots 12 \qquad (3)$$

Spośród wielu innych możliwości najprostszą będzie hipoteza liniowej zależności między w i a

$$w_i = p_i\, a + q_i \qquad i = 1, 2, \ldots 12 \qquad (4)$$

Hipotezą tą zajmiemy się obecnie szerzej. Dla krótkości będziemy mówili o sezonowości liniowej.

Jak wyznaczyć wartość parametrów p i q? Oznaczając $P_i = 1 + p_i$, postulujemy[3])

$$s^2 = \frac{1}{12\,n}\left[\sum_1^{12} \sum\nolimits_i (x - P_i\, a - q_i)^2\right]. \qquad (6)$$

gdzie Σ_i oznacza sumę wartości dla i — tego miesiąca ze wszystkich n lat zbadanych. Gdyby wartości P_i i q_i dla $i = 1, 2, \ldots 12$ były między sobą niezależne, rozwiązanie zagadnienia otrzymania minimum na (6) byłoby zupełnie banalne. Osiągnięte przy tym założeniu wartości parametrów oznaczymy przez P' i q'; przedstawiają się one następująco:

$$P'_k = \frac{\sum_k ax - \frac{1}{n}\sum_k a \sum_k x}{\sum_k a^2 - \frac{1}{n}\left(\sum_k a\right)^2} \qquad (7)$$

$$q'_k = \frac{1}{n}\left(\sum\nolimits_k x - P' \sum\nolimits_k a\right) \qquad (8)$$

W rzeczywistości jednak musimy postulować pewne związki między parametrami P i q. Zakładamy mianowicie, że jeśli na przestrzeni roku wartość normalna zmiennej pozostaje bez zmiany, to suma składników sezonowych dla poszczególnych miesięcy powinna się równać zeru, czyli

$$\sum_1^{12}(P_i\, a + q_i) = 12a \qquad (9)$$

dla każdego a, co znowu sprowadza się do dwóch warunków:

$$\sum_1^{12} P_i = 12 \qquad (10)$$

$$\sum_1^{12} q_i = 0. \qquad (11)$$

Pragniemy więc sprowadzić do minimum następujące wyrażenie:

$$s^2 = \frac{1}{12n}\left\{\sum_1^{11} \sum\nolimits_i (x - P_i a - q_i)^2 + \sum\nolimits_{12}\left[x - \left(12 - \sum_1^{11} P_i\right) a + \sum_1^{11} q_i\right]^2\right\} \qquad (12)$$

---

[1]) Wg naszej pracy „Sezonowość liniowa", Studia Ekonomiczne III, 1936, str. 65—68.

[2]) Jako wartość tę będziemy przyjmowali w praktyce średnią ruchomą 13-miesięczną, w której miesiące skrajne ważone są po ½.

[3]) Por. J. Wiśniewski „O mierze skuteczności eliminowania sezonowości", Prace Instytutu Badania Koniunktur Gospodarczych i Cen zesz. 1 z r. 1935, str. 9—15.

Dla każdego $i = k$ otrzymujemy parę równań normalnych, a mianowicie

$$\sum_k ax - P_k \sum_k a^2 - q_k \sum_k a - \sum_{12} ax + \\ + \left(12 - \sum_1^{11} P_i\right) \sum_{12} a^2 - \left(\sum_1^{11} q_i\right) \sum_{12} a = 0 \tag{13}$$

oraz

$$\sum_k x - P_k \sum_k a - nq_k - \sum_{12} x + \\ + \left(12 - \sum P_i\right) \sum_{12} a - n \sum_1^{11} q_i = 0. \tag{14}$$

A zatem mamy dwadzieścia dwa równania normalne i tyleż niewiadomych (P_{12} i q_{12} określamy z (10) i (11)) i zagadnienie jest w teorii rozwiązane. Jednakże rachunki praktyczne są ogromnie żmudne, tak iż trzeba się uciec do metod przybliżonych. Określenie parametrów rozłożymy na dwa etapy.[1])

Ze wzorów (7) i (8) okazuje się, że dla danego k wartość P' nie zależy od wartości q i od wyboru osi współrzędnych, gdy tymczasem wartość q zależy zarówno od P jak i od położenia początku układu. Wobec tego określimy najpierw P_k Zamiast (12) sprowadzimy do minimum

$$s'^2 = \frac{1}{12n} \left\{ \sum_1^{11} \sum_i (x' - P_i a')^2 + \\ + \sum_{12} \left[x' - \left(12 - \sum_1^{11} P_i\right) a' \right]^2 \right\} \tag{15}$$

gdzie

$$x' = x - \frac{1}{n} \Sigma_i x, \quad a' = a - \frac{1}{n} \Sigma_i a.$$

Różniczkując względem P_k $(k \pm 1)$, otrzymujemy jedno z jedenastu równań normalnych, mianowicie

$$\sum_k a'x' - P_k \sum_k a'^2 - \sum_{12} a'x' + 12 \sum_{12} a'^2 - \\ - \left(\sum_1^{11} P_i\right) \left(\sum_{12} a'^2\right) = 0 \tag{16}$$

Łatwo zauważyć, że

$$P_k = \frac{P_1 \Sigma_1 a'^2 - \Sigma_1 a'x' + \Sigma_k a'x'}{\Sigma_k a'^2}. \tag{17}$$

Podstawiając w równanie normalne dla P_1, dostajemy

$$P_1 = \frac{\Sigma_1 a'x'}{\Sigma_1 a'^2} + \frac{12 - \sum_1^{12} \frac{\Sigma_i a'x'}{\Sigma_i a'^2}}{\sum_1^{12} \frac{1}{\Sigma_i a'^2}} \cdot \frac{1}{\Sigma_1 a'^2} \\ = P'_1 + \frac{12 - \sum_1^{12} P_i'}{\sum_1^{12} \frac{1}{\Sigma_1 a'^2}} \cdot \frac{1}{\Sigma_1 a'^2} \tag{18}$$

i analogicznie dla innych P_k $(k \pm 1)$.

Teraz oznaczymy q. Sprowadza się to do zagadnienia sezonowości stałej addytywnej, przy czym jednak zamiast a rolę wartości normalnych będą spełniały wartości Pa. Zagadnienie to zostało rozwiązane w naszym przyczynku w Pracach Instytutu Badania Koniunktur Gospodarczych i Cen, zesz. 1 z r. 1935 (wzór 12). Otrzymujemy

$$q_k = \frac{1}{n}\left(\sum_k x - P_k \sum_k a\right) - \\ - \frac{1}{12n} \sum_1^{12} \left(\sum_i x - P_i \sum_i a\right) \tag{19}$$

Eliminowanie sezonowości odbywa się w ramach hipotezy liniowej przez zastosowanie wzoru

$$\bar{a} = \frac{x - q}{P} \tag{20}$$

gdzie $\bar{a}$ oznacza wartość „normalną" wyliczoną z równania (20). Nasuwa się tutaj pewna subtelna kwestia: czy dla eliminowania sezonowości nie należałoby stosować innych równań, niż dla opisu sezonowości, podobnie jak inne jest z reguły równanie regresji X względem Y niż Y względem X. Sądzimy, że na pytanie to należy odpowiedzieć przecząco, ze względów ściśle związanych z naturą szeregów rozwojowych. Poszczególne obserwacje nie są losowane kolejno dla najrozmaitszych wartości a, lecz przeciwnie dla wartości a niezbyt różniących się między sobą (choć może nam nieznanych). Wskutek tego w obrębie roku następuje (w przybliżeniu) wyrównanie odchyleń wartości x od ich równań

[1]) Por. J. Wiśniewski „A problem in least squares" The Annals of Mathematical Statistics zesz. 3 z r. 1937, str. 145—148.

regresji względem a, gdy tymczasem podobne wyrównanie odchyleń wartości a od ich równań regresji względem x bynajmniej nie następuje i przez to dla poszczególnych lat mielibyśmy odczytane z równań regresji wartości a systematycznie za niskie lub za wysokie. Możnaby wprawdzie zaradzić temu przez postawienie postulatu jak w (9), lecz wprowadzenie odpowiednich poprawek prawie zidentyfikowałoby oba typy linii regresji, nie mówiąc o wielkich trudnościach rachunkowych.

Powyżej opisana metoda stosowana jest przez Instytut Badania Koniunktur Gospodarczych i Cen w tych wypadkach, gdzie prostsze metody nie dają zadowalających wyników.

Jako przykład zastosowania opisanej powyżej metody weźmiemy liczby zatrudnienia (tysiące robotniko - godzin na 1 dzień roboczy) w przemyśle mineralnym. Obliczamy najpierw średnie ruchome. Najpraktyczniej odbywa się to w ten sposób: przypuśćmy, że rozporządzamy danymi od lipca 1930 r. począwszy.

Rok i miesiąc		Tys. rob.-godz.		
1930	VII	412,0		
	VIII	411,7		
	IX	397,0		
	X	352,3		
	XI	283,8		
	XII	220,9	3470,2	
1931	I	198,2		3392,6
	II	195,9		
	III	199,4		
	IV	209,4		
	V	271,9		
	VI	317,7		
	VII	334,4		
	.	.		
	.	.		
	.	.		
	.	.		

Dodajemy najpierw wszystkie liczby od VII. 1930 r. do VI. 1931 r. włącznie: suma ich wynosi 3470,2. Następnie znajdujemy takąż sumę dla okresu od VIII. 1930 do VII. 1931 włącznie; nie potrzebujemy jednak dodawać wszystkich dwunastu składników, wystarczy wykonać następujące działanie: 3470,2—412,0+334,4 = 3392,6. Odjęliśmy liczbę z VII. 1930 i dodaliśmy liczbę z VIII. 1931. Postępując nadal w podobny sposób, znajdujemy sumy dla wszystkich kolejnych okresów 12-miesięcznych. Sumy te dodajemy po dwie: pierwszą z drugą, drugą z trzecią i t. d. W naszym przykładzie znajdujemy 3470,2 + 3392,6 = 6862,8. Dzieląc tę ostatnią liczbę przez 24, otrzymujemy wartość średniej ruchomej dla I. 1931. 6862,8 : 24 = 286,0.

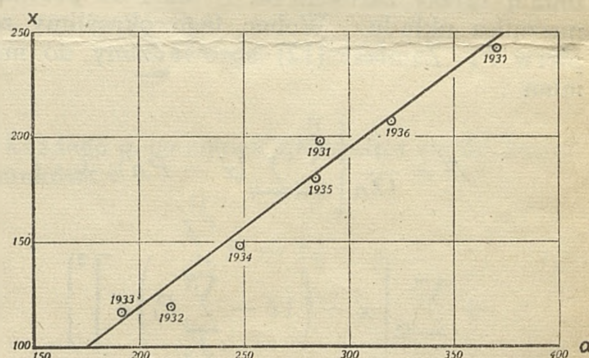
Znalazłszy w ten sposób średnie ruchome dla wszystkich miesięcy okresu badanego, robimy oddzielnie dla stycznia ze wszystkich lat, oddzielnie dla lutego itd., zestawienia liczb surowych (x) i średnich ruchomych (a).

Podajemy takie zestawienie dla stycznia:

Rok	x	a	a^2	ax
1931	198,2	286,0	81796	56685
1932	118,6	214,9	46182	25487
1933	116,5	192,2	36941	22391
1934	148,2	248,0	61504	36753
1935	180,2	284,0	80656	51177
1936	207,5	319,5	102080	66296
1937	241,7	370,1	136974	89453
Razem	1210,9	1914,7	546133	348242
Średnia	173,0	273,5	78019	49749

Wartości a i x zestawiamy na wykresie 1.

WYKRES 1.



Zatrudnienie w przemyśle mineralnym w styczniu.
x — liczby surowe, a — średnie ruchome.

Już pobieżny rzut oka na ten wykres przekonywa nas, że punkty wyznaczone przez pary wartości a i x dla poszczególnych lat leżą w przybliżeniu na linii prostej, wobec czego zastosowanie metody liniowej jest uzasadnione.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Razem
F'	0,756	0,736	0,911	1,177	1,252	1,071	1,030	1,057	1,157	1,114	0,972	0,898	12,131
v	311	294	277	258	245	235	350	340	344	344	336	329	3663
d	— 0,011	— 0,011	— 0,010	— 0,009	— 0,009	— 0,008	— 0,013	— 0,012	— 0,012	— 0,012	— 0,012	— 0,012	— 0,131
P	0,745	0,725	0,901	1,168	1,243	1,063	1,017	1,045	1,145	1,102	0,960	0,886	12,000
q'	—30,8	—26,6	—53,1	—81,4	—35,7	47,1	68,9	68,8	30,7	11,6	13,4	—24,3	—14,4
q	—29,6	—25,4	—51,9	—80,2	—34,5	48,3	70,1	70,0	31,9	12,8	14,6	—26,1	0,0

Celem obliczenia wartości P'_1 ze wzoru (7), który przepiszemy w poniższej formie

$$P'_1 = \frac{\frac{1}{n}\Sigma_1 ax - \frac{1}{n}\Sigma_1 a \cdot \frac{1}{n}\Sigma_1 x}{\frac{1}{n}\Sigma_1 a^2 - \left(\frac{1}{n}\Sigma_1 a\right)^2} \qquad (7')$$

musimy jeszcze wyznaczyć wartości ax i a^2 dla poszczególnych lat oraz ich sumy; dzieląc przez $n = 7$, otrzymujemy średnie dla okresu badanego (p. tabliczka powyżej). Podstawiamy teraz we wzór (7')

$$P'_1 = \frac{49749 - 273{,}5 \cdot 173{,}0}{78019 - (273{,}5)^2} = \frac{49749 - 47316}{78019 - 74802} =$$

$$= \frac{2433}{3217} = 0{,}756$$

Podobne obliczenia przeprowadzamy dla pozostałych miesięcy, od lutego do grudnia, i zestawiamy je w poniższej tabliczce.

Suma współczynników P' dla wszystkich miesięcy wynosi 12,131, jest zatem większa od postulowanej przez nas liczby 12 (p. wzór 10). Wobec tego ostateczne współczynniki P będą mniejsze od współczynników P'. Wzór (18) przepiszemy w ten sposób:

$$P_k = P'_k + d_k \qquad (18')$$

gdzie

$$d_k = \left(12 - \sum_1^{12} P'_i\right) \frac{v_k}{v_i}$$

a z kolei

$$v_k = \frac{10^6}{\frac{1}{n}\Sigma_k a^2 - \left(\frac{1}{n}\Sigma_k a\right)^2}$$

Liczbę 10^6 wprowadzamy w liczniku wyrażenia na v_k, aby uniknąć ułamków.

Wypełniamy teraz w tabliczce wiersz zatytułowany v. Dla stycznia otrzymujemy

$$v_I = \frac{10^6}{3217} = 311.$$

Wyliczenie d — zgodnie z naszymi wzorami — sprowadza się do „rozdzielenia" (z odwrotnym znakiem) nadwyżki (12,131 — 12) na poszczególne miesiące proporcjonalnie do wartości v. W szczególności dla stycznia mamy

$$d_I = -0{,}131 \cdot \frac{311}{3663} = -0{,}011$$

$$P_I = 0{,}756 - 0{,}011 = 0{,}745.$$

Obliczenie q jest znacznie mniej mozolne od obliczenia P. Wzór (19) napiszemy w ten sposób:

$$q_k = q'_k - e \qquad (19')$$

gdzie

$$q'_k = \frac{1}{n}\Sigma_k x - \frac{P_k}{n}\Sigma_k a$$

oraz

$$e = \frac{1}{12}\sum_1^{12} q_i$$

Dla stycznia

$$q'_I = 173{,}0 - 0{,}745 . 273{,}5 = 173{,}0 - 203{,}8 = -30{,}8$$

Suma wartości q' dla wszystkich miesięcy równa się — 14,4, a więc

$$e = -1{,}2; \quad q_I = -30{,}8 + 1{,}2 = -29{,}6$$

T. zw. eliminowanie sezonowości odbywa się przez podstawienie odpowiednich liczb do wzoru (20), który wygodniej jest przedstawić w następującej formie:

$$\bar{a} = \alpha_k x + \beta_k \qquad (20')$$

gdzie $\bar{a}$ oznacza wartość z wyeliminowaniem sezonowości metodą liniową, zaś

$$\alpha_k = \frac{1}{P_k}; \quad \beta_k = -\frac{q_k}{P_k}$$

Oczywiście równanie (20') ma dla każdego miesiąca inne współczynniki. Np. dla stycznia znajdziemy

$$\alpha_I = \frac{1}{0{,}745} = 1{,}342$$

$$\beta_I = \frac{29{,}6}{0{,}745} = 39{,}7$$

Dla roku 1936 np. wyliczymy:

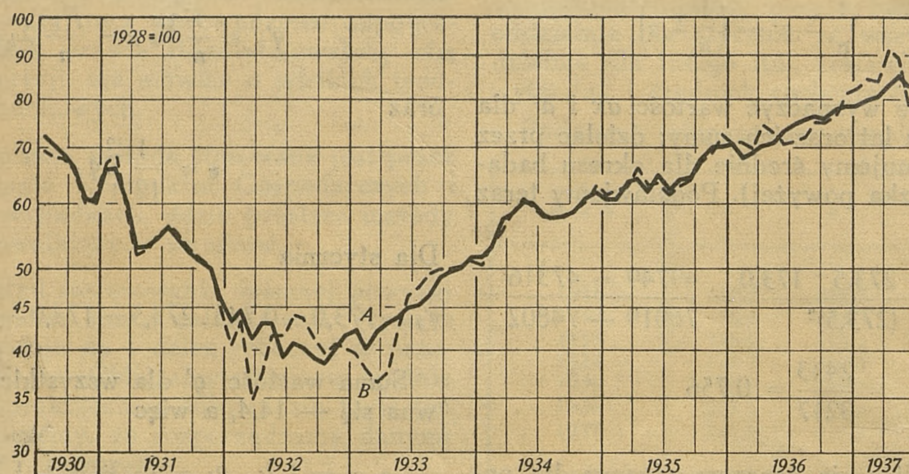
$$\bar{a} = 134{,}2 . 207{,}5 + 39{,}7 = 278{,}5 + 39{,}7 = 318{,}2.$$

Na wykresie 1 odpowiednikiem równania (20') jest prosta. Pozwala nam ona na znalezienie wartości $\bar{a}$ metodą graficzną, bez wyliczenia. Odczytujemy po prostu wartość a, odpowiadającą na prostej zaobserwowanej wartości x.

Na wykresie 2 przedstawiony jest przebieg wskaźników zatrudnienia w przemyśle mineralnym z wyeliminowaniem sezonowości: A — opisaną powyżej metodą „liniową", B — metodą sezonowości stałej. Jakkolwiek warunki porów-

nania są korzystne dla sezonowości stałej, gdyż współczynniki obliczono na podstawie całego okresu zbadanego (VII. 1930 — VI. 1937), to jednak porównanie to wypada na korzyść sezonowości liniowej, zwłaszcza w tych latach, gdy zachodziły większe wahania (1932, 1933, 1937).

WYKRES 2.



Zatrudnienie w przemyśle mineralnym. Wskaźniki z wyeliminowaniem sezonowości: *A* — metodą liniową, *B* — metodą stałych współczynników.



CENA 1 ZŁOTY **Druk ukończono 4 marca 1939 r.**

Adres redakcji i administracji: Warszawa, Al. Jerozolimskie 85. — Telefon sekretariatu redakcji 7.24.82, — telefon administracji 7.28.01.

All communications should be addressed: Instytut Badania Koniunktur Gospodarczych i Cen, Warszawa, Al. Jerozolimskie 85.

Druk. „KADRA" Warszawa, Długa 50